chinaXiv:201705.00728v1

# STKOS 中领域本体模型框架研究*

马雨萌 [1]　刘凤红 [2]　黄金霞 [1]
[1]（中国科学院文献情报中心　北京　100190）
[2]（中国科学院植物研究所　北京　100093）

**摘要：**［**目的/意义**］构建普适的领域本体模型框架和规范的方法体系，为 STKOS 实现动态建模的本体工具集和面向领域应用的本体网络提供建设依据。［**方法/过程**］调研分析不同层次的领域本体模型，总结能够为 STKOS 本体建设提供的参考，构建 STKOS 中学科领域本体模型的三层架构，提出各层框架可遵循的方法，并以中国科学院植物研究所的需求为例，阐述了本文模型框架在构建植物多样性领域本体的应用。［**结果/结论**］按照领域本体模型三层架构所构建的植物多样性领域本体，较之其他领域本体，具有知识组织体系清晰、构建过程高效、所支撑应用场景丰富的优点，因此，该研究为 STKOS 快速而有效地构建领域本体提供了实用框架，未来可应用于知识服务平台语义层的建设。

**关键词：**领域本体　STKOS　模型框架　植物多样性领域
**分类号：**G250

## 1 引言

科技文献信息的知识组织体系（Scientific & Technological Knowledge Organization Systems, STKOS）建设促进了科技文献信息服务向知识服务的转变，为海量科技文献资源的整合组织、深层揭示和知识关联提供支撑[1]。领域本体作为知识组织体系中规范描述和语义组织学科领域核心知识的模型，通过对科技文献信息资源深层知识内容的集成、挖掘和关联，能够支持面向科学研究的学科化服务和应用。由于在构建过程中缺乏统一清晰的框架和规范完善的依据，目前已有领域本体的交互和集成效果较弱，STKOS 涉及学科众多，为了形成可复用的本体资源，各学科领域本体需要提供规范化的构建场景和步骤依据，因此，建立统一的领域本体模型框架和方法体系，能够为 STKOS 本体建设实现动态建模的本体工具集和面向领域应用的本体网络提供基础。

本文研究了可供各学科参考的领域本体模型框架，提出学科领域本体在构建中可依据的过程和方法，为开发领域本体动态建模的本体工具集提供场景和流程依据，实现 STKOS 领域本体网络的构建。因此，本文在对不同层次领域本体进行调研的基础上，提出学科领域本体模型的三层架构，并以植物多样性领域为例，阐述了本文模型框架在构建领域本体中的具体应用。

## 2　不同层次领域本体模型分析

构建领域本体的目标就是要捕获相关领域的知识，提供对该领域知识的共同理解，确定该领域内共同认可的术语，并从不同层次的形式化模式上给出这些术语之间相互关系的明确定义。根据概念是否面向学科领域及其描述层次，目前领域本体主要从无具体学科特征的顶层本体（Top Ontology）、表达学科共有特征的领域上层本体（Top-domain Ontology）、面向具体建设目标的领域应用本体（Domain Ontology）方面进行研究[2]。

### 2.1 顶层本体模型

顶层本体是适于所有学科领域的概念模式，为各学科构建领域本体提供可重用的普适框架，易于适应不同领域需求的变化，能够增强各学科领域本体之间的互操作性。目前关于顶层本体构建的研究，通过直接分析现实世界中基本的类和关系，或从不同的抽象角度建立展

---

* 本文系国家科技支撑计划项目“面向外文科技文献信息的知识组织体系建设与应用示范（STKOS）”子课题“面向外文科技文献信息的超级科技词表和本体建设”（课题编号：2011BAH10B01）研究成果之一。
[作者简介]马雨萌（ORCID：0000-0003-3593-114X），中国科学院文献情报中心助理馆员，E-mail:mayumeng@mail.las.ac.cn；刘凤红（ORCID：0000-0002-3633-1464），中国科学院植物研究所图书馆高级工程师；黄金霞（ORCID：0000-0001-8705-0067），中国科学院文献情报中心副研究馆员。

现和表示现实世界的事物框架，典型本体模型有 BFO（Basic Formal Ontology）和 DOLCE（Descriptive Ontology for Linguistic and Cognitive Engineering），为不同学科领域提供了资源组织和描述的可扩展模型。

BFO 本体是由形式化本体和医学信息科学研究所（Institute for Formal Ontology and Medical Information Science，IFOMIS）构建的描述现实世界存在事物的基本框架，BFO 通过对现实领域中类和关系的抽象，以多元化的角度展现了现实世界中存在的实体，能够为不同领域本体的构建提供通用的上层框架。BFO 首先依据在时间范围内实体存在的方式，将现实世界存在的事物分成两类，持续性实体和过程、活动等临时性实体[3]。对于持续存在的实体，既有独立存在的物质实体以及空间区域等非物质实体，也包括需要物质实体作为载体而存在的依赖实体，例如固有特征、角色、功能。BFO 框架为不同领域知识关系的构建提供了上层关系体系，梳理了现实世界中实体之间的基本关系。对于持续性实体类体系，主要为物质实体之间的部分整体关系、固有特征与物质实体的本质即属关系、物质实体与所处地理空间的位置关系。临时性实体用于表达持续性实体的时间信息，两大类实体之间的关系大致分为物质实体与所在过程的参与关系、物质实体与相关时间范围的存在关系，过程与所在地点的事件发生关系。BFO 本体框架提供了对现实世界抽象的基本实体概念和基本关系，为不同学科领域本体的构建提供了可扩展的顶层本体，目前 Gene Ontology、Protein Ontology 以及 Ontology for Biomedical Investigations 均采用了 BFO 来支持相关研究领域中实体类型体系和关系的构建[4]。

DOLCE 本体作为 WonderWeb 项目基础本体库中最先开发的模块，以通用顶层本体的形式明确表示了与本体库中各扩展模块的关系，以及用于发现已有本体或词汇资源的深层知识结构。DOLCE 的持久性实体类和临时性实体类与 BFO 类型体系所表达的含义类似，同时，DOLCE 将实体的特征值、属性、时空范围等具体描述信息在抽象类（Abstract）体系中组织。DOLCE 本体应用于 WordNet 概念的规范化组织和描述，由于 WordNet 缺乏将现实知识和实体规范概念化的组织框架，DOLCE 本体通过引入概念之间的继承关系、个体和概念之间的实例化关系，与 WordNet 的上层概念进行比较和映射，解决了 WordNet 中概念和实例混淆的局面；另外，WordNet 将时间、空间等对象概念和属性、关系等描述概念混杂为抽象概念，然而 DOLCE 界定了对象概念和描述概念的范畴，分别为其构建了相应的类型体系，WordNet 的这些概念可在相应体系中组织管理[5]。

通过对 BFO 类和关系体系以及 DOLCE 在描述和组织知识中应用的分析，顶层本体对物质实体的描述大体从固有特征、空间及时间等方面，将这些能够表达实体基本特征的信息分别抽象为相应的上层类，并通过基本实体关系的捕获，为规范化组织概念提供了上层关系体系，因此该类本体能够反映现实世界中一般实体的特征状态和存在关系。

**2.2 领域应用本体**

领域应用本体一般面向学科中的细分领域，依据明确的领域需求和应用目标，构建能够较为全面描述该领域知识的本体模型。本文选取材料学和生物医学领域，分析领域应用本体在这两个学科中的研究进展。

目前材料科学领域本体的构建研究描述了材料科学的知识概念和关系，在数据集成、文本知识抽取、跨学科分类模式构建以及数据库之间的信息交互等方面得到应用。Ashino 和 Fujita 构建的本体面向材料选择的过程，以标准化的模式实现了相关领域数据资源和材料属性的集成[6]。PLINIUS 本体提供了陶瓷材料领域的知识模型，基于该领域的原子概念和推理规则构建较为复杂的概念，形成能够用于陶瓷材料领域文本知识抽取的知识体系[7]。Tanaka 基于纳米领域的标准和权威资源构建本体，提供了与纳米科学技术相关的跨学科分类模式[8]。这些领域应用本体分别面向材料科学的具体领域，提供了梳理其应用领域中知识及其关系的模型。

chinaXiv:201705.00728v1

为了对生物医学领域海量的数据和事实知识进行有效的组织管理和分析利用，OBO 本体库（Open Biological and Biomedical Ontologies）通过构建一系列的领域本体，描述各细分领域中的基础实体[9]。基因本体（Gene Ontology）是其中的典型案例，通过构建可共享和结构化术语模型的方式用于标注有机体分子特征，以分子功能（molecular function）、生物过程（biological process）、细胞成分（cellular component）三个本体分支提供了描述有机体内基因产物的结构化词汇集合，着重于表达基因产物特征和详情，以及构建这些生物学领域术语之间的明确关系。基因本体用于基因产物资源的有效组织及其生物学特征的准确标注，从而支持跨库资源的查询、检索和集成[10]。

虽然这些领域应用本体作为相关领域的知识模型，较为清晰地描述了领域的核心概念和知识关系，但是这些本体只关注各自局限的应用领域，即使在同一学科中，由于缺乏概念集成和相互关联的机制，领域本体的组织体系和结构框架不一致，较难满足本体可扩展性的需求，阻碍了知识管理任务中本体的语义标准化和互操作。

**2.3 领域上层本体**

为了解决领域应用本体在可扩展和语义交互方面存在的上述问题，领域上层本体捕获了用于描述学科领域基本特征的核心类和关系，作为面向该学科所有研究领域的普适本体，提供规范化组织和描述领域概念及其关系的上层框架。本部分仍以生物医学和材料学领域为例，分析这两个学科领域上层本体的相关研究。

为了解决生物医学不同领域本体的语义标准化和互操作问题，OBO 构建了一个通用的领域上层本体，用于实现生物医学领域本体的集成和标准化。一方面根据 Gene Ontology、Cell Ontology、ChEBI（Chemical Entities of Biological Interest），分析得到能够体现生物医学领域共有特征的上层类和关系[11]；另一方面扩展和再设计原有的 GENIA 框架，解决了 GENIA 类型描述或组织结构存在的缺乏标准化的问题，构建 BioTop（Biological Top-Level）本体作为集成各生物医学领域本体的框架[12]。为了涵盖生物化学领域的概念，BioTop 中的部分类迁移至新建本体 ChemTop，BioTop 包含了对分子生物和生物医学领域基本类型的描述，ChemTop 则展现了生物化学领域的关键概念类型，二者共同作为领域上层本体，能够较为全面地定义和描述生物医学领域中的基本实体和关系。图 1 表达了 BioTop 和 ChemTop 作为领域描述的中间层，与顶层本体和领域应用本体的无缝衔接。BioTop 和 ChemTop 类的构建基于顶层本体 BFO，BioTop 继承了 BFO 的 Material Entity 并按照实体粒度扩展子类体系，关系的表达则来源于顶层关系本体 RO（Relation Ontology）的关系类型[2]。对于具体应用领域的本体，Gene Ontology、Cell Ontology 与 BioTop，以及 ChEBI 与 ChemTop 分别通过类的匹配和映射，实现了不同 OBO 领域应用本体的集成。BioTop 和 ChemTop 为不同来源的生物医学领域本体提供了标准化和体系化的语义集成框架，其定义的概念类型成为关联 OBO 领域本体的桥梁，能够增强信息抽取和文本挖掘等自然语言处理应用的效果。

chinaXiv:201705.00728v1

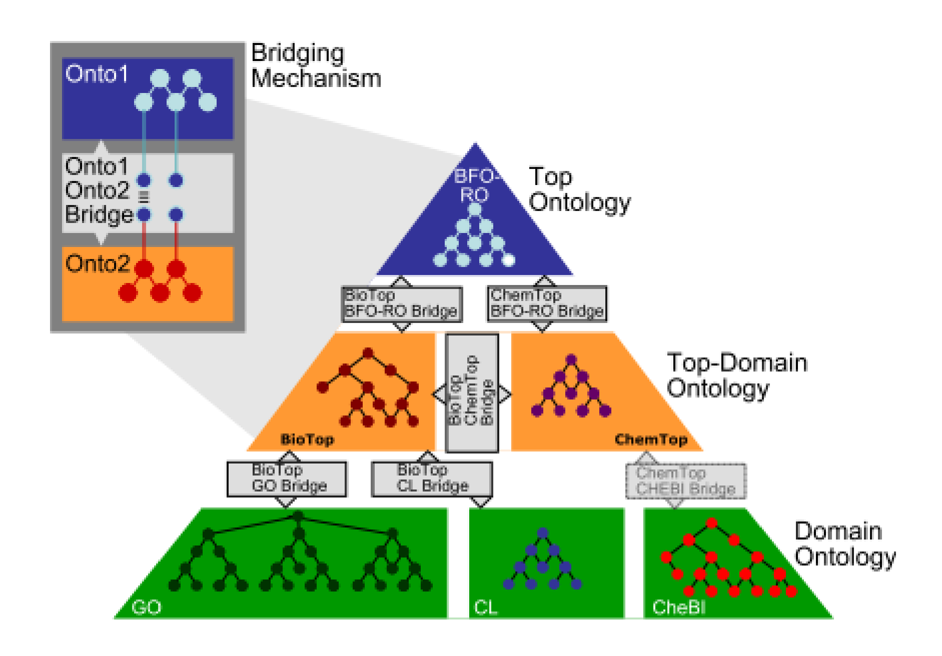



图 1 BioTop 和 ChemTop 作为领域上层本体的模型图[2]

在材料科学领域本体的研究中，MatOnto 本体通过构建材料科学领域的通用、可扩展模型，促进材料科学各领域数据资源和实验过程的交互、重用和集成。MatOnto 本体的建设目标在于构建材料学领域上层概念类型的组织框架，所遵循的原则首先基于顶层本体构建，且其知识组织框架能够提供丰富、易于理解的抽象概念类型以及清晰的组织结构；另外依据已有的领域本体或权威词汇表，尽可能保证本体的概念体系符合领域大多研究人员的认知需求。MatOnto 本体以顶层本体 DOLCE 作为上层基础框架，从 DOLCE 的三个大类 Endurants、Perdurants、Abstract 出发，以扩展其子类的方式构建类体系。MatOnto 复用了 Ontolingua 本体的标准单位类、W3C 的时间本体、语义网研究社群本体（SWRC）等本体或分类表，通过扩展 EXPO 本体的事件和过程概念来丰富科学实验的描述。在扩展材料科学领域概念方面，通过引入本体或词汇表扩展了 MatOnto 上层类，例如，与学科分类词表 JACS（Joint Academic Classification of Subjects）中材料学科的关联实现对材料学概念体系的表达，从特征、分类体系、生产测量过程、材料结构和关联数据等方面实现了对材料学领域核心概念的全面表达[13]。MatOnto 作为材料领域的上层本体，提供了组织管理材料学核心知识以及集成相关领域本体的基础框架，基于本体模型中通用的材料领域术语集合，实现了异构和不同数据源的数据发现、检索、集成。

领域上层本体 BioTop、 ChemTop 和 MatOnto 作为领域本体框架的中间层，从 BFO 等描述现实世界普适概念的顶层本体扩展而来，捕获了具有学科领域特征的上层类和关系，实现了领域普适顶层本体和领域应用本体之间类和关系的衔接，能够促进学科中其他领域本体的集成以及新细分领域本体的构建。

通过分析以上本体模型的类型体系、关系框架、建模过程、构建方法、应用实践，并根据 STKOS 的需求和目标，这些不同层次领域本体能够为 STKOS 本体建设提供以下参考：

（1）STKOS 面向多学科建立领域本体网络和动态建模的本体工具集，因此，需要统一的框架体系对各领域进行概念的规范整合和关系的清晰梳理。顶层本体描述了各学科领域普适的基本实体概念和知识关系，作为上层框架能够通过类和关系的扩展，将各学科领域的知识及其关系纳入规范化的组织结构中。

（2）由于学科涉及众多细分领域，STKOS 领域本体网络实现本体重用、映射和关联的目标，要求所构建领域本体具有可扩展性且彼此能够语义交互。领域上层本体作为某一学科

chinaXiv:201705.00728v1

领域的上层基础框架，规范一致的扩展结构促进了本体的集成、复用和互操作，可用于新细分领域的本体构建或已有本体结构的优化，而且对体现学科领域共有特征核心概念和知识关系较为全面的捕获也能够保证学科的知识需求。

（3）STKOS 目标通过本体工具集的建立实现领域本体的动态建模，需要明确的本体构建场景和流程框架。在领域上层本体的构建过程中，继承顶层本体的类和关系体系、集成和复用已有领域本体或权威词汇表等活动，为开发支持本体工程情境化的工具提供了规范的场景依据。此外，领域应用本体可基于顶层本体和领域上层本体的模型框架，通过从抽象到具体的类和关系扩展，实现动态和快速建模。

## 3 STKOS 领域本体框架及其构建方法

本文构建了面向 STKOS 本体建设的学科领域本体模型框架，由顶层本体、领域上层本体、领域应用本体组成的建立领域本体的三层架构，如图 2 所示，并提出领域本体在构建中可依据的过程和方法，以期为 STKOS 各应用领域构建本体和本体工具集开发所基于的情境化流程提供参考。











图 2 STKOS 领域本体模型框架和构建方法体系

**（1） 顶层本体**

上层框架中的实体不具有学科特征，一般从特征、空间、时间等方面对现实世界中物质进行描述。不同学科可根据其领域中实体的特征，选择 BFO、DOLCE 等认可度较高的顶层本体作为学科概念类型扩展的基础框架，上层关系框架可参考顶层关系本体 RO 的基本关系。顶层本体要具有适于广泛领域进行概念扩展和知识组织的基本实体及其关系体系，能够为学科领域概念的定义和组织提供普适框架。

**（2） 领域上层本体**

作为连接顶层本体和具体领域应用本体的中间框架，领域上层本体面向某一学科由顶层本体扩展而来，为学科各应用领域的本体构建或集成提供可扩展的模型。由于 STKOS 的超级科技词表对于一些学科领域规范整理了科技概念以及继承梳理了关系属性[1]，因此一方面基于超级科技词表，制定词表向本体转化的映射规则；另一方面调研分析领域中成熟本体、权威词汇表以及新建知识组织体系作为补充资源，通过词表概念的遴选、规范和本体的重用、重构、映射，以及继承词表或已有本体的关系属性，形成能够体现学科共有特征的核心概念

chinaXiv:201705.00728v1

及其关系体系，从而构建适于该学科组织和描述知识的上层本体。

（3） 领域应用本体

以上两层结构为底层领域应用本体的构建提供了扩展依据和基础框架，根据领域应用需求和具体建设目标，分析从哪些方面能够较为全面地描述该领域中的关键知识，从而明确如何扩展领域上层本体，形成描述该具体应用领域中资源的知识模型。

## 4 植物多样性领域本体构建示例

本文以植物多样性领域为例，针对中国科学院植物研究所有效组织植物多样性领域重点关注的物种信息的需求，包括物种的名称、分类位置、形态描述以及分布地信息。本文按照提出的三层框架以及建设方法体系，示例植物多样性领域本体的构建过程，图 3 展示了植物多样性领域本体的知识组织体系，虚线方框、椭圆框、实线方框分别代表了顶层本体、生物多样性领域上层本体、底层植物多样性领域扩展本体的类，捕获了各层级的核心知识关系，展示了不同层级架构之间的扩展体系，实现植物多样性领域核心概念的组织和关键信息源的描述。限于篇幅，图 3 只展示了多样性领域的部分类。

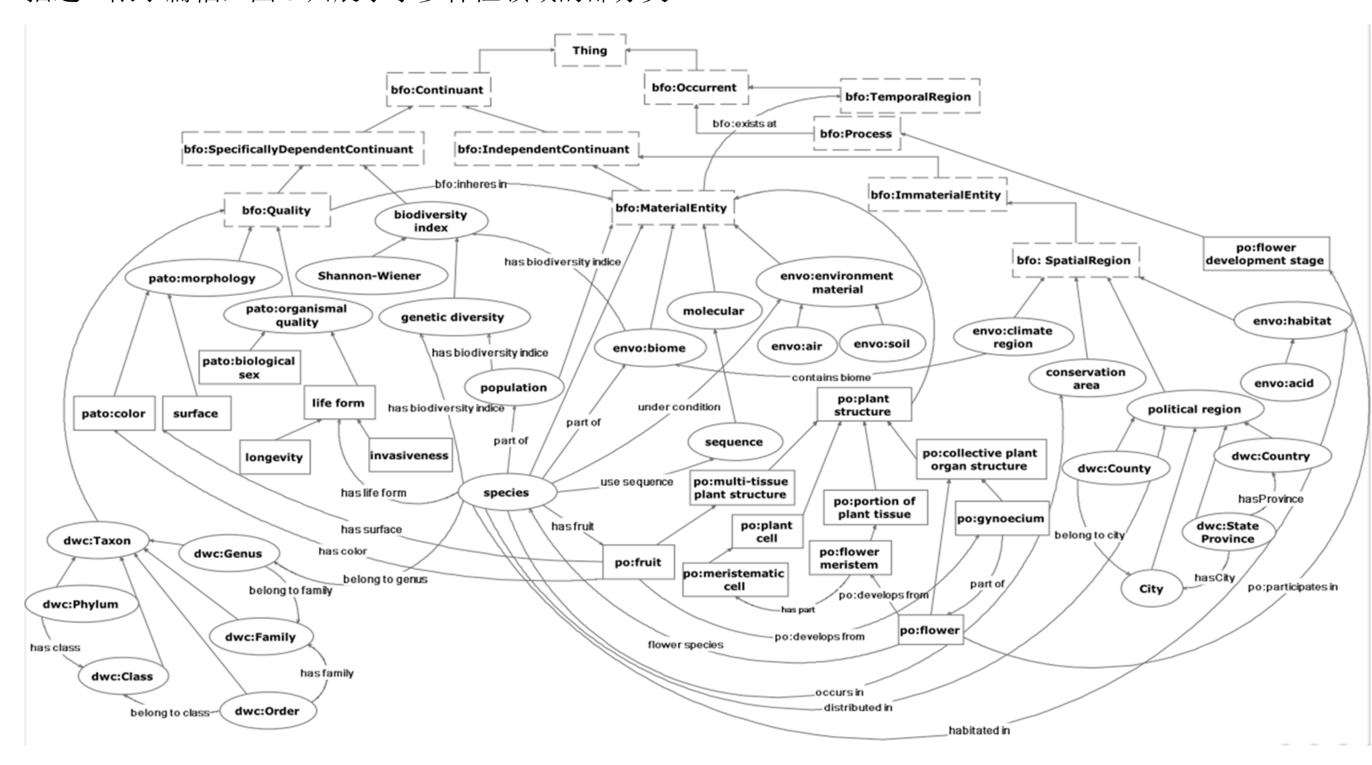



图 3 植物多样性领域本体三层框架结构

### 4.1 顶层本体框架

由于 BFO 本体在 OBO 本体库建设中的成功应用，显示了其能够较好地适用于生物学不同领域本体的构建[2]，因此，本研究通过复用 BFO 本体构建顶层本体的类和关系。顶层本体需要为学科领域提供普适框架，其类和关系应能够表示学科领域中实体的一般状态，按照 BFO 本体的“物质（Material）-特征（Quality）”模型框架，同时考虑时空、过程等因素，构建生物学领域表达核心概念的顶层知识组织体系。

### 4.2 生物多样性领域上层本体框架

本研究的中层框架面向生物多样性领域，由顶层本体的类和关系扩展而来，首先梳理、遴选、分类整理了 STKOS 关于生物多样性、环境物质等领域词表的核心概念，依据词表向本体转化的映射规则，并参考复用 Darwin Core[14]、表型和特征本体（Phenotype And Trait Ontology，PATO）[15]、环境本体（Environment Ontology, ENVO）[16]等领域本体或词表，从描述物种、生态系统、遗传的多样性方面，构建描述生物多样性核心概念及其关系的领域上

chinaXiv:201705.00728v1

层框架。

**（1）描述物种多样性的知识组织体系**

生物多样性领域的组织框架以物种（species）为分界线和核心，类体系捕获了物种的分类术语、形态特征及其所处的生境、栖息地、保护区、行政区域和环境条件，建立了分类概念之间、物种与特征、物种与地域、物种与环境实体的知识关系，从而能够揭示物种的分类位置、物种的基本特征、地域的物种清单、物种依赖的环境物质等信息。

有关生物有机体名称和分类的术语概念界定，通过复用 Darwin Core 的相关元素描述了门、纲、目、科、属、种的概念[17]，并建立了相邻分类位置概念之间的互逆关系，用于描述物种所在的分类位置信息。针对生物个体及其组成器官的特征，依据 PATO 本体扩展了形态特征（morphology）、有机体特征（organismal quality）等类，提供了物种通用物质特征的上层框架。生物多样性有关地域信息的表达，包括物种所处（located in）的气候区（climate region）、出现（occurs in）的保护区（conservation area）、栖息（habitated in）的生境（habitat）以及分布（distributed in）的行政区域（political region），从而可获得对物种地域信息的全方位描述，以及梳理某一地域中的物种清单。此外，构建了表达环境物质的类体系，包括大气（air）、土壤（soil）等物种生存所依赖（under condition）的环境因素。

**（2）描述生态系统多样性的知识组织体系**

此部分表达了可度量种群、生态系统层级的多样性，不仅揭示了物种与种群、生态系统的部分整体关系（part of），而且构建了丰富的指标类，用于定量化表征生态系统的多样性，以及反映了地域环境的复杂性对生物多样性的影响。

生态系统（biome）下包含水生群系（aquatic biome）和陆生群系（terrestrial biome）及其下属的各个群系类型，例如水生群系下又划分为淡水生态系统和海洋生态系统。本体构建了定量表征生物多样性的指标类体系（biodiversity index），这些主要用于描述和衡量生态系统多样性大小和格局的具体指数，包括描述物种和遗传的不同层级的多样性指标，例如常用的 Shannon-Wiener 指数，也包括一些最近兴起的表征功能多样性和系统发育多样性的指标。另外，以上对于物种相关地域环境的描述，同样适用于建立生态系统类与上述地域环境类的关系，表达某一生态系统所处的地域信息以及所需的环境条件。

**（3）描述遗传多样性的知识组织体系**

为了描述物种的遗传多样性，本体建立了物种与基因序列（sequence）的关联，用于描述物种使用的基因组片段，并且通过为物种赋予数值属性（sequence data），提供物种在 NCBI 上的物种 ID，可以进行关联和检索 DNA 条码测序数据。

### 4.3 植物多样性领域的本体扩展

由于植物多样性是生物多样性研究的细分领域，因此通过为生物多样性领域上层本体扩展子类体系的方式，能够基本满足植物多样性描述的需求。由生物多样性领域上层框架，同样地将遴选的 STKOS 植物领域词表概念进行转化，还主要复用了 PO 本体（Plant Ontology）[18]的类和关系，从植物的组成结构、生长发育、个体及其组成部分的特征、生活型信息等方面，扩展了本体中面向植物多样性领域的概念和关系。

本体构建了叶、花、果实等植物组成结构类，并捕获了这些类之间的发育关系（develops from），从形状（shape）、结构（structure）、颜色（color）、表面（surface）等方面，扩展了植物个体和结构层面的特征[19]，同时也重点捕获了植物生长发育过程涉及的细胞、组织等微观概念，从而揭示了处于某个生长过程或发育阶段的某一物种及其组成结构的特征。如图 3 所示，雌蕊（gynoecium）是花（flower）的组成结构，果实（fruit）由雌蕊发育而来；花在（participates in）发育阶段（flower development stage）中，由花分生组织（flower meristem）发育而来，这些分生组织由分生细胞（meristematic cell）组成。此外，生活型信息（life form）类体系用于表达植物丰富的生活特征，例如，根据植物的寿命（longevity），生活型包括一

chinaXiv:201705.00728v1

chinaXiv:201705.00728v1

年生、二年生、多年生。

同时，对于植物个体或组成的部分形态特征，可采用添加数值属性的方式描述，例如叶片的长度（length）、宽度（width）等特征。另外，丰富了本体中物种、地域、分类、形态等类在植物多样性领域的实例，并补充了植物物种的同义词（含同物异名、缩写词、英文名）。

### 4.4 本体的效果评价

本文按照领域本体模型三层架构所构建的植物多样性领域本体，较之其他领域本体，具有如下优点：

（1）清晰的知识组织体系

由本文模型框架形成的此领域本体具有逐层深入的组织体系，依据各层级的知识组织特征和构建流程，顶层本体具有适于生物学知识组织的 MQ 模型框架，领域上层本体满足了生物多样性领域知识从物种多样性、生态系统多样性、遗传多样性方面进行表征的需求，底层本体在此框架的基础上针对植物领域深入扩展。因此，该本体的类及其关系体系较为清晰地反映了植物多样性领域的知识脉络。

（2）更为高效的构建过程

本文提出的模型框架，为领域本体的构建提供了可扩展的组织架构和规范的建设场景，对于构建流程和方法的梳理也促进了领域本体建设效率的提高。领域上层框架明确了 STKOS 词表向本体转化的规则，以及生物多样性领域的知识表达需求，为细分领域本体的建设提供了快速扩展的依据。

（3）支撑更加丰富的应用场景

该领域本体对植物多样性领域中核心概念和知识关系的全方位捕获，能够支持丰富的应用场景。一方面实现知识的多维度语义标注，通过标引地区的物种、生态系统类型及其多样性指数的数值，可实现该地区多样性的表征；再比如，某一地区或生态系统的环境条件及其与生物多样性的关系，可通过标引地区、生态系统类型、环境条件和生物多样性指数等实现该类问题的回答。另一方面促进领域数据的组织和知识的揭示，可依据植物多样性领域本体中的植物分类，将物种进行属、种等层次的组织，以及可梳理某一地区（如自然保护区）包含的物种数目、清单；在揭示领域知识方面，利用植物多样性领域本体中的植物分布地域、生境类型、生物多样性指标、形态特征、生长发育期等类和实例标注来源数据，能够识别与汇集有关植物分布、生境、多样性指标、形态、生长发育阶段对应的数据，通过揭示这些领域知识，可用于描述植物在不同尺度下生长发育和形态可塑性的趋异性。

根据以上效果分析，此本体基本涵盖了植物多样性领域的核心概念，能够满足植物所组织和揭示该领域重点关注信息的需求。由于词表的结构各异，可参考复用的领域本体面向不同的建设需求，因此，虽然本文提出了领域本体的普适框架，但是在应用过程中，也会遇到难以用统一的映射规则对词表概念进行转化，以及不同来源概念如何准确归类的问题。另外，生物多样性领域框架限定了植物多样性领域概念的扩展来源，底层本体只能由已有的上层类来添加子类体系，因此，对于从多个来源词表遴选的植物多样性领域概念，会存在少数概念较难找到合适上层类的情况。

## 5 结语

目前领域本体建设工作在实践中常遇到诸多困难，尤其在针对具体学科领域各异的应用需求时。STKOS 项目为解决这个问题提供了契机，面向以本体支持知识服务的需求，研究建立领域本体模型，利用项目建设的大量词表和概念网络，面向特定学科领域和应用场景，快速而有效地完成领域本体的构建，正如在植物多样性领域本体建设中的工作。未来基于该领域本体在植物多样性领域知识服务平台语义层的应用，将实现知识导航、语义内容关联、自动聚类等功能，为用户提供语义层次、一体化植物多样性信息资源服务，其可行性将基于 STKOS 中领域本体建设和应用进一步完善。

chinaXiv:201705.00728v1

参考文献：

[1]孙坦,刘峥.面向外文科技文献信息的知识组织体系建设思路[J]. 图书与情报,2013(1):2-7.

[2]Stenzhorn H, Beißwanger E, Hoek L, et al.BioTop and ChemTop–Top-Domain Ontologies for Biology and Chemistry[EB/OL].[2014-05-15].http://ceur-ws.org/Vol-401/iswc2008pd_submission_36.pdf.

[3]Smith B.On Classifying Material Entities in Basic Formal Ontology[EB/OL].[2014-05-06].http://ontology.buffalo.edu/smith/articles/Material_Entities.pdf.

[4]Arp R, Smith B.Function, Role, and Disposition in Basic Formal Ontology[EB/OL].[2014-05-06].http://precedings.nature.com/documents/1941/version/1/files/npre20081941-1.pdf?origin=publication_detail.

[5]Gangemi A, Guarino N, Masolo C, et al.Sweetening Ontologies with DOLCE[C].In:Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web: Springer-Verlag.2002:166-181.

[6]Ashino T,Fujita M.Definition of a web ontology for design-oriented material selection[J].Data Science Journal,2006,5:52-63.

[7]Der P V, E.Vet P, Speel P-h, et al.The PLINIUS ontology of ceramic materials[C].In:the Eleventh European Conference on Artificial Intelligence (ECAI'94) Workshop on Comparison of Implemented Ontologies, Amsterdam, The Netherlands: Springer-Verlag.1994:8-12.

[8]Tanaka M.Toward a Proposed Ontology for Nanoscience[C].In:CAIS/ACSI 2005:Data, Information and Knowledge in a Networked World, The University of Western Ontario, London.2005.

[9]Smith B, Ashburner M, Rosse C, et al.The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration[J/OL].Nature Biotechnology,2007(25)[2014-05-15]. http://www.nature.com/nbt/journal/v25/n11/full/nbt1346.html.

[10]The Gene Ontology Consortium.Creating the Gene Ontology Resource: Design and Implementation[EB/OL].[2014-05-21].http://www.geneontology.org/meeting/AnnotationCampReading2006/GO-GenRes-2001.pdf.

[11]Beisswanger E, Schulz S, Stenzhorn H, et al.BIOTOP: An Upper Domain Ontology for the Life Sciences[EB/OL].[http://www.imbi.uni-freiburg.de/ontology/biotop/publications/ao08.pdf.

[12]Schulz S, Beisswanger E, Hahn U, et al.From GENIA to BIOTOP-Towards a Top-Level Ontology for Biology[EB/OL].[2014-05-15].http://www.imbi.uni-freiburg.de/ontology/biotop/publications/fois06.pdf.

[13]Cheung K, Drennan J, Hunter J.Towards an Ontology for Data-driven Discovery of New Materials[EB/OL].[2014-05-21].http://www.itee.uq.edu.au/eresearch/papers/2008/MatOnto_Cheung_Hunter_final.pdf.

[14]Wieczorek J, Döring M, De Giovanni R, et al.Darwin Core[EB/OL].[2014-06-10].http://rs.tdwg.org/dwc/.

[15]Mungall C J, Gkoutos G V, Smith C L, et al.Integrating phenotype ontologies across multiple species[J].Genome Biol,2010,11(1):R2.

[16]Morrison N, Ashburner M, Field D, et al.The environment ontology: Linking environmental data[C].In:Proceedings of the European conference: Towards Environment. Opportunities of SEIS and SISE: Integrating Environmental Knowledge in Europe, Masaryk University, Brno, Czech Republic.2009:606-608.

[17]Wieczorek J, Bloom D, Guralnick R, et al.Darwin Core: An Evolving Community-Developed Biodiversity Data Standard[J/OL].PLoS ONE,2012,7(1)[2014-06-10]. http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0029715.

[18]Jaiswal P, Avraham S, Ilic K, et al.Plant Ontology (PO): a Controlled Vocabulary of Plant Structures and

Growth Stages[J].Comp Funct Genomics,2005,6(7-8):388-97.
[19]Andrew R D, Matthew J Y,James P B.Time to change how we describe biodiversity[J].Cell,2012,27(2):78-84.

**作者贡献声明：**
马雨萌：设计研究方案，文献调研，实施研究过程，构建示例本体，起草、撰写、修订论文；
刘凤红：示例本体的概念遴选和应用评价；
黄金霞：提出研究方向，指导示例本体构建。

Research on Model for Domain Ontology of STKOS

Ma Yumeng[1]  Liu Fenghong[2]  Huang Jinxia[1]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] Institute of Botany, Chinese Academy of Sciences, Beijing 100093

**Abstract: [Purpose/Significance]** This study aims to build a general model for domain ontology and normative method system, in order to provide STKOS with structure basis for the dynamic modeling of ontology toolkit and the network construction of domain ontology. **[Method/Process]** Based on the research of domain ontology of different levels from which practical conclusions for STKOS are summarized, the paper builds model of three-tier architecture for domain ontology and puts forward to methods that construction of each tier can follow. Finally, taking the institute of botany for example, the paper elaborates the application of this model in the construction of domain ontology of plant diversity. **[Result/Conclusion]** Compared with other domain ontology, the domain ontology of plant diversity according to three-tier architecture has a series of outstanding features, such as clear knowledge organization system, efficient build process and support for rich usage scenario. Therefore, this study provides STKOS with a practical framework for quickly constructing effective domain ontology, which can be applied to knowledge service platform for the construction of the semantic layer.

**[Keywords]** domain ontology  STKOS  architecture of model  plant diversity

chinaXiv:201705.00728v1